Area-specific predictions of unwanted events using multivariate modeling of water data

Dominiak, Katarina Nielsen; Hindsborg, Jeff; Pedersen, Lene Juul; Kristensen, Anders Ringgaard

Publication date:
2018

Document version
Publisher's PDF, also known as Version of record

Document license:
Unspecified

Citation for published version (APA):
Dominiak, K. N., Hindsborg, J., Pedersen, L. J., & Kristensen, A. R. (2018). *Area-specific predictions of unwanted events using multivariate modeling of water data*. 27-30. Paper presented at 2nd International Conference on Agro Big Data and Decision Support Systems in Agriculture, Lleida, Spain.

Area-specific predictions of unwanted events using multivariate modeling of water data

Katarina N. Dominiak[1], Jeff Hindsborg[1], Lene J. Pedersen[2], Anders R. Kristensen[1]

[1] University of Copenhagen
Grønnegårdsvej 2, 1870 Frederiksberg C, Denmark
knd@sund.ku.dk, ark@sund.ku.dk

[2] Aarhus University
Blichers Allé 20, 8830 Tjele, Denmark
lene.juulpedersen@anis.au.dk

Abstract

In this study simultaneously monitored drinking patterns from multiple pens in one weaner section and one finisher section herd are modeled in a multivariate *dynamic linear model* (DLM). Correlations between the monitored data are incorporated in the model allowing it to include interactions between pens in each section. A *two-sided tabular CUSUM* is applied as control chart, whereby alarms for outbreaks of unwanted events amongst the pigs can be generated for specific pens in the herd. Area-specific alarms enable the manager to include specific knowledge of the animals in the targeted area and choose the best suited intervention.

1 Introduction

The primary cause of excess false alarms from sensor-based early warning systems for livestock production is difficulties at achieving detection performances so high that the number of false alarms will be acceptable in a real-life production herd, as discussed by [5, 7]. Therefore more focus should be aimed at developing modeling methods or post-processing methods which rank, sort, or prioritize alarms before they are communicated to the manager.

A new approach for prioritizing alarms from early warning systems is to relate alarms to a specific area of the herd using a spatial detection model. This would allow the manager to include any specific knowledge of the animals in the targeted areas and choose the better intervention for preventing an outbreak of an unwanted event or for reducing the consequences of one.

Modern Danish production herds for growing pigs (weaners 7 - 30 kg and finishers 30 - 110 kg) are very well suited for spatial modeling due to both the construction of the herds and the highly systematic managerial routines. The herds consist of identical sections which consist of a number of identical pens. Furthermore the pigs are inserted in the sections following a strategy where pigs of the same age are inserted in a section on the same day, and the section is emptied and disinfected before the insertion of a new batch of pigs [2].

The objective of this paper is to validate a spatial model, which detects systematic changes in the water consumption of growing pigs throughout the growth period. Area-specific alarms are generated by a control chart, and they will act as decision support in the daily management of the herd.

2 Materials and methods

2.1 Data

The model was originally developed on water consumption data from two different herds (Herd A, a finisher herd. Herd B, a weaner herd) where it was able to model simultaneously monitored data from multiple pens in multiple sections of a herd. When applied it could identify a specific pen or a specific section in a herd where an unwanted event was going to occur [4, 6].

In the current paper, the model is validated on data from a single weaner section and a single finisher section in an independent herd (Herd C). Data sets from weaners and finisher are modelled separately.

The water data was monitored using photo-electric flow sensors (RS V8189 15mm Diameter Pipe) [1], which were installed in eight pens (four double pens) in a weaner section, and in sixteen pens (eight double pens) in a finisher section in the same herd. Data from Herd C was monitored while another project was conducted in the herd. Therefore 50% of the pens in a section contained pigs with intact tails and 50% of the pens contained pigs with docked tails. In every other batch, pigs with intact tails were inserted in pens which contained pigs with docked tails in the previous batch and vice versa.

Model validation was conducted following two strategies of internal and external evaluation. In *Int. evaluation* the first four batches of Herd C were defined as training data for the model, whereas the last two batches of Herd C were defined as test data and used for performance evaluation. In *Ext. evaluation* data from Herd A was used as training data for evaluation of all six batches from the finisher section of Herd C, and data from Herd B was used as training data for evaluation of all six batches from the weaner section of Herd C.

Every morning, the caretaker registered events of diarrhoea and tail biting in each pen. These event registrations constitute the *gold standard* when evaluating model performance. Evaluation was conducted for the prediction of either of the two events separately and for both events together.

2.2 General model

The water consumption over time was modelled simultaneously for all sensors in each section using a multivariate DLM as described by [10]:
The observation vector $Y_t = (Y_{1t}, \dots, Y_{nt})'$, contains the observation at time t for each of the n sensors. Both the relation between Y_t and the underlying parameter vector θ_t at time t, and the evolution of the system over time, are described through an observation equation and a system equation (Equations (1) and (2) respectively):

Observation equation

$$Y_t = \mathbf{F}_t' \boldsymbol{\theta}_t + \nu_t, \quad \nu_t \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_t), \tag{1}$$

and

System equation

$$\theta_t = \mathbf{G}_t' \theta_{t-1} + w_t, \quad w_t \sim \mathcal{N}(\mathbf{0}, \mathbf{W}_t). \tag{2}$$

The overall aim of the DLM is to predict the next observation of the monitored variable by estimating the parameter vectors, $\theta_1, \dots, \theta_t$, from the observations, $Y_1, \dots, Y_t$. Every observation is added to the model's prior knowledge of the modeled system, and this dynamic updating enables the model to predict the next observation with increased certainty over time. When a new observation is made, the predicted value and the observed value are compared, and any differences between the predicted and the actual observations are due to the two error terms, ν_t and w_t [3].

If the pigs in the pens follow their normal drinking pattern and drink the expected amounts of water, the prediction of the next observation is close to perfect, and any prediction error will be small. If, on the other hand, something cause the pigs to drink more or less than expected, the prediction error will be larger. A systematic change in the normal drinking pattern will generate a sequence of larger prediction errors, which will be reflected as an alarm generated by a *CUSUM*.

2.3 Modeling diurnal patterns

A previous study by [8] found, that three harmonic waves of lengths 24h, 12h, and 8h describe the diurnal pattern, which characterize the water consumption of growing pigs. Each harmonic wave is expressed as cyclic models in a dynamic linear model through the trigonometric *Fourier form representation of seasonality* [10] as:

$$\mathbf{F}_t^h = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{G}_t^h = \begin{pmatrix} \cos(\omega) & \sin(\omega) \\ -\sin(\omega) & \cos(\omega) \end{pmatrix}. \tag{3}$$

with $\omega = 2\pi/24$ yielding a wave with a period of 24h, $\omega = 2\pi/12$ a wave with a period of 12h, and $\omega = 2\pi/8$ a wave with a period of 8h.

However, since pigs drink more water as they grow, a trend must be added to the diurnal pattern in order to model the full drinking pattern. Therefore a dynamic *linear growth model* was used for modeling the underlying level of water consumed as well as the increase in the level from time $t-1$ to t. It is described by [10] as:

$$\mathbf{F}_t^l = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{G}_t^l = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \tag{4}$$

2.4 Area-specific alarms

Detection of alarms and irregular drinking patterns was done using Tabular CUSUM control chart as described by [9]. The Cusum generated cumulated sums of the positive and the negative forecast errors separately over time, and plotted them as *Upper Cusum* and *Lower Cusum* respectively. If either of the Cusums exceeded a defined threshold the monitored process was considered out of control, and an alarm was generated.

A measure of test accuracy, the *Area Under Curve* (AUC) was calculated in order to evaluate the model performance. This was done separately for the two sections. The detection model was evaluated for the ability to predict events in a specific pen in the section by comparing the generated alarms with the *gold standard* registrations from the personnel. Time windows applied began two days before the event and included the day of the event as well.

3 Results

The accuracies for predicting an event in a specific pen were high following both evaluation strategies (see Table 1). The prediction of tail bites separately obtained the highest performance for both sections and evaluation strategies, whereas the prediction of diarrhea separately performed the poorest. Some differences in the AUC can be seen between the two evaluation strategies. However, the relatively small differences implies that variance estimates generated on data from one herd can be successfully applied to data from other herds.

	Weaners		Finishers	
Event	Int. evaluation	Ext. evaluation	Int. evaluation	Ext. evaluation
Tail bites	0.92	0.88	0.84	0.83
Diarrhea	0.80	0.77	0.72	0.71
Both events	0.81	0.81	0.73	0.76

Table 1: AUC (area under curve) for prediction of either tail biting, diarrhea or both events in specific pens in a section of finishers and a section of weaners following two evaluation strategies; Int. evaluation and Ext. evaluation

4 Conclusion and perspectives

The detection accuracies obtained in this study implies that changes in drinking patters are highly informative for the prediction of tail bites, and to some extend outbreaks of diarrhea amongst growing pigs. The findings furthermore supports the results found in [4] which implies that the alarms are very accurate in pointing out a specific area where unwanted events are about to occur. This means that when a manager receives an alarm for a specific area in the herd, it is very likely that an event is going to occur in that area within the next two or three days. Due to the systematic managerial routines in the herd, the manager knows the age and size of pigs in any section at any time, as well as high-risk periods for outbreaks of specific events. This means that an area-specific alarm can provide a valuable support for the manager in deciding both which pen or section to attend first and which intervention to implement. Although the full potential of the model is better exploited when applied to data from multiple pens in multiple sections, the present study clearly shows that the model is robust and can be applied to a wide span of data sets.

5 Acknowledgements

This research was funded by the Danish Council for Strategic Research (The PigIT project, Grant number 11-116191). We further wish to thank the anonymous farmers, the Danish Pig Research Centre and the technical staff (particularly Mads Ravn Jensen) at Aarhus University for installing and supervising the sensors and for taking care of the daily observations in the herd.

References

[1] Anonymous. Instruction leaflet flow sensors - 15 mm dia. pipe v8189, 2000.

[2] R. D. A. Cameron. A review of the industialisation of pig production worlwide with particular reference to the asian region. *Food and Agriculture Organization of the United Nations*, 2000.

[3] K. N. Dominiak. *Spatial modeling of drinking patterns as a tool for reducing alarms in pig production*. PhD thesis, The Graduate School of Health and Medical Sciences, University of Copenhagen, 2017.

[4] K. N. Dominiak, Hindsborg J., L. J. Pedersen, and A. R. Kristensen. Spatial modeling of pigs' drinking patterns as an alarm reducing method. II. Application of a multivariate dynamic linear model. *Submitted for journal publication*, 2018.

[5] K. N. Dominiak and A. R. Kristensen. Prioritizing alarms from sensor-based detection models in livestock production - a review on model performance and alarm reducing methods. *Computers and Electronics in Agriculture*, 133:46–67, 2017.

[6] K. N. Dominiak, L. J. Pedersen, and A. R. Kristensen. Spatial modeling of pigs' drinking patterns as an alarm reducing method. I. Developing a multivariate dynamic linear model. *Accepted for journal publication*, 2018.

[7] H. Hogeveen, C. Kamphuis, W. Steeneveld, and H. Mollenhorst. Sensors and clinical mastitis - the quest for the perfect alert. *Sensors*, 10(9):7991–8009, 2010.

[8] T. N. Madsen, S. Andersen, and A. R. Kristensen. Modelling the drinking patterns of young pigs using a state space model. *Computers and Electronics in Agriculture*, 48(1):39–62, 2005.

[9] Douglas C. Montgomery. *Statistical Quality Control A Modern Introduction*. John Wiley & Sons, Inc, seventh edition, 2013.

[10] Mike West and Jeff Harrison. *Bayesian Forecasting and Dynamic Models*. Springer, 2nd edition, 1999.